**Airbnb Dynamic Pricing Recommendation Report**

**1. Objective**

The objective of this project is to develop a dynamic pricing recommendation engine for Airbnb listings. By analyzing historical data, the model aims to predict optimal listing prices based on various factors such as location, season, amenities, reviews, and property attributes.

**2. Dataset Overview**

* **Source**: Airbnb listings data (CSV format)
* **Target Variable**: price
* **Key Features**:
  + property\_type, room\_type, accommodates, bedrooms, bathrooms
  + review\_scores\_rating, number\_of\_reviews
  + amenities, availability\_365
* **Size and Scope**: Includes multiple listings across cities and countries

**3. Exploratory Data Analysis (EDA)**

* The dataset contained several null values; columns with excessive missing data were dropped or filled as appropriate.
* Listings with prices above $1000 were treated as outliers and removed.
* Price distributions varied significantly by city and room type.
* Correlation analysis indicated that:
  + accommodates, bedrooms, and bathrooms are strongly correlated with price
  + review\_scores\_rating has a mild positive correlation
  + The newly created amenity\_count feature showed moderate correlation with price

**4. Feature Engineering**

**4.1 Amenity Count**

The amenities column, originally a string of items within curly braces, was converted into a list. Then a new column amenity\_count was created to quantify the number of amenities provided in each listing.

df['amenities'] = df['amenities'].str.replace('[{}"]', '', regex=True).str.split(',')

df['amenity\_count'] = df['amenities'].apply(lambda x: len(x) if isinstance(x, list) else 0)

**4.2 Column Cleanup**

* Non-numeric columns such as name, host\_id, city, cancellation\_policy, etc., were removed or one-hot encoded.
* Only numeric columns were retained for modeling using:

X = df\_model.drop('price', axis=1).select\_dtypes(include=['number'])

**5. Modeling Approach**

* **Algorithm Used**: Linear Regression (via sklearn)
* **Target Variable**: price
* **Train/Test Split**: 80% training, 20% testing
* **Preprocessing**:
  + Removed non-numeric columns
  + Filled or dropped null values
* **Model Evaluation**:
  + Performance measured using R² Score
  + Example Score: Approximately 0.61 (depends on final cleaned dataset)

**6. Key Predictive Features**

The following features had the strongest influence on price:

* accommodates
* bedrooms
* bathrooms
* review\_scores\_rating
* amenity\_count

These factors are crucial when suggesting dynamic pricing.

**7. Pricing Recommendations**

* Listings with more bedrooms, bathrooms, and higher accommodates values should be priced higher.
* A higher review\_scores\_rating justifies a premium.
* Listings with more amenities can command better prices.
* Adjust prices based on seasonal availability or booking trends (if such data is available).

**Prepared by:**  
B. Anitha Pai  
[Project: Elevate Labs Airbnb Pricing Engine]